DOI:
10.1007/s10670-015-9749-7

Document Version
Peer reviewed version

Citation for published version (APA):
Fumagalli, R. (2016). Why We Cannot Learn from Minimal Models. *ERKENNTNIS*, 433-455.
https://doi.org/10.1007/s10670-015-9749-7

Why We Cannot Learn from Minimal Models

Abstract

Philosophers of science have developed several accounts of how consideration of scientific models can prompt learning about real-world targets. In recent years, various authors advocated the thesis that consideration of so-called minimal models can prompt learning about such targets. In this paper, I draw on the philosophical literature on scientific modelling and on widely cited illustrations from economics and biology to argue that this thesis fails to withstand scrutiny. More specifically, I criticize leading proponents of such thesis for failing to explicate in virtue of what properties or features minimal models supposedly prompt learning and for substantially overstating the epistemic import of minimal models. I then examine and refute several arguments one may put forward to demonstrate that consideration of minimal models can prompt learning about real-world targets. In doing so, I illustrate the implications of my critique for the wider debate on the epistemology of scientific modelling.

Keywords: Scientific Modelling; Minimal Models; Representation; Learning; Epistemology of Modelling.

Introduction

Philosophers of science have developed a number of accounts of how consideration of scientific models can prompt learning about real-world targets (see e.g. Cartwright, 1999 and 2009, Mäki, 2005 and 2009, Morgan, 1999 and 2001, and Sugden, 2000 and 2009). These accounts offer dissimilar indications as to what conditions scientific models must satisfy to prompt learning about such targets. Two clusters of accounts are usefully contrasted in this context. On the one hand, some focus on the diverse representational functions served by models and allege that consideration of scientific models can prompt learning about real-world targets provided that these models satisfy specific criteria of representational adequacy (see e.g. van Fraassen, 1980, and French, 2003, on isomorphism; Giere, 2004, Teller, 2001, and Weisberg, 2012, on similarity; and Mäki, 1992 and 2009, on partial resemblance). On the other hand, others emphasize models' suitability to serve as surrogate systems (see e.g. Suarez, 2004, and Swoyer, 1991) and maintain that highly idealized scientific models can prompt learning by providing means of conceptual exploration (see e.g. Hausman, 1992, ch.5) and credible 'parallel worlds' (see e.g. Sugden, 2000 and 2009) that do not aim to accurately represent any real-world target.

These two clusters of accounts are not incompatible (see e.g. Mäki, 2009, for an account of models as isolations and surrogate systems), but offer dissimilar indications as to what conditions scientific models must satisfy to prompt learning about real-world targets. In spite of these differences, all those accounts rest on the presupposition that consideration of scientific models can prompt learning about real-world targets only if some world-linking relations (e.g. similarity, isomorphism, resemblance) hold between such models and targets. As Morgan puts it, "if we want to use [a model] to learn about the world, the model needs to map onto the real world" (1999, 366; see also Mäki, 1992 and 2009, and Sugden, 2000 and 2013, for similar remarks).

In recent years, various authors (e.g. Grüne-Yanoff, 2009 and 2013a, and Knuuttila, 2009) have called this widespread presupposition into question. In their view, modellers can learn about real-world targets by consideration of so-called *minimal models*, i.e. models that "lack any similarity, isomorphism or resemblance relation to the world, [are] unconstrained by natural laws or structural identity, and [do not] isolate any real factors" (Grüne-Yanoff, 2009, 83). Consideration of minimal models is said to prompt *learning* in the sense that constructing, analysing and manipulating these models can *justifiably* "affect one's confidence in necessity or impossibility hypotheses" about real-world targets without imposing any similarity, isomorphism, resemblance, etc. requirement between such models and targets (Grüne-Yanoff, 2009, 81). Let us call the thesis that we can learn from minimal models in this sense 'LMM'.

If correct, LMM would have far-reaching implications for scientific modellers, since necessity and impossibility hypotheses are advocated, discussed and tested in a vast range of disciplines (see e.g. Casini, 2014, and Knuuttila, 2009, on economic models; see also Batterman and Rice, 2014, and Rohwer and Rice, 2013, on biological models). In this paper, I examine prominent calls in support of LMM and argue that such calls are vulnerable to severe objections. I then draw on often-quoted examples of putative learning from minimal models to demonstrate that LMM itself substantially overstates the epistemic import of such models. More specifically, I shall argue that *contra* LMM, consideration of minimal models cannot prompt learning about real-world targets unless one supplements these models with additional information or presuppositions concerning such targets.

My critique of LMM, which qualifies and extends a preliminary work on the epistemology of minimal models (Fumagalli, 2015), proceeds as follows. In *Section 1*, I outline recent calls in support of LMM and criticize the proponents of LMM for failing to specify in virtue of *what properties* or *features* minimal models can supposedly prompt learning about real-world targets. In *Section 2*, I draw on widely cited illustrations from economics and biology to argue that consideration of minimal models cannot *per se* prompt learning about such targets. In particular, I aim to demonstrate that LMM falls prey to the following dilemma. On the one hand, truly minimal models - i.e. models that are minimal in the sense indicated by LMM - lack the evidential and epistemic resources to prompt learning about real-world targets. On the other hand, several models that *prima facie* seem minimal in the sense indicated by LMM can prompt learning about real-world targets, but succeed in doing so only if supplemented with additional information or presuppositions about such targets. In *Section 3*, I examine and refute three arguments one may put forward to defend LMM against my critique. More specifically, I shall consider in turn the argument from how-possibly explanations, the argument from clusters of models, and the argument from heuristic applications. In doing so, I illustrate how my critique of LMM bears on the wider debate on the epistemology of scientific modelling.

I shall expand in the following sections on the main reasons why I think that LMM fails to withstand scrutiny. For now, it suffices to note that my critique of LMM is not intended to suggest - and by no means implies - that building and manipulating minimal models is a futile modelling exercise. In particular, my critique allows that consideration of minimal models may provide epistemically informative insights concerning the possible worlds posited by these models. Still, it challenges the proponents of LMM to both specify in virtue of what properties or features minimal models can prompt learning about real-world targets and identify cases where consideration of these models prompts such learning.[1]

The expression 'minimal model' has been employed in dissimilar senses by the philosophers and the practitioners of different disciplines (e.g. ecology, set theory, theoretical physics). Two uses of 'minimal model' besides the one adopted by Grüne-Yanoff (2009) are popular among scientific modellers. Some employ this expression to indicate models that include only "those factors that make a difference to the occurrence and essential character of [the investigated phenomena]" (Weisberg, 2007a, 642). Others, instead, speak of 'minimal models' to refer to caricatures that capture only the purportedly universal and repeatable features of their targets (see e.g. Batterman, 2002, and Goldenfeld, 1992) and may be used to show why targets "that are heterogeneous at smaller scales […] all display the same large-scale behavior" (Batterman and Rice, 2014, 349-350).

Below I shall employ the expression 'minimal model' in the sense indicated at p.2 unless specified otherwise. Moreover, I shall devote particular attention to Grüne-Yanoff's (2009) defence of LMM, since I regard it as the most detailed and widely cited case for this thesis (see e.g. Cartwright, 2009, de Donato Rodriguez and Zamora Bonilla, 2009, and Sugden, 2009 and 2013). Still, I shall expand in various passages on how my critique of LMM bears on other calls for the possibility of learning from minimal models, including calls that draw on less restrictive characterizations of 'minimal model' than the characterization underlying LMM (see e.g. the discussion in *Section 1*; see also footnote no.11, on Rohwer and Rice, 2013, and footnote no.13, on Batterman and Rice, 2014).[2]

[1] I employ the expression 'possible worlds' broadly to indicate the hypothetical, imaginary, or counterfactual scenarios and states of affairs that are posited by specific models. My broad use of this expression covers what other authors (e.g. Morgan, 2001, and Sugden, 2000) have called 'model worlds' and 'worlds of the model'. Also, I speak of 'properties' and 'features' - as opposed to 'property' and 'feature' - to allow for the possibility that distinct properties and features enable learning in different modelling contexts. I shall comment on this context-dependency and its implications for the merits of LMM in *Section 1*.

[2] I shall occasionally state that minimal models 'are assumed to lack any similarity, isomorphism, resemblance, etc. relation' to real-world targets as an abbreviation for the characterization of 'minimal models' reported in the text. The proponents of LMM (e.g. Grüne-Yanoff, 2009, 84) occasionally hint at a broader characterization of 'minimal model', according to which the set of minimal models includes both models that actually lack any similarity, isomorphism, resemblance, etc. relation to real-world targets and models for which no relation of similarity, isomorphism, resemblance, etc. to specific real-world targets has been determined by modellers. I shall expand in *Section 1* on the differences between distinct characterizations of 'minimal model' and the implications of such differences for the merits of LMM. For now, I rely on the narrower characterization reported in the text both because the proponents of LMM repeatedly employ this characterization to single out minimal models

1. In virtue of *what properties* can minimal models prompt learning?

LMM states that consideration of minimal models can prompt learning in the sense that constructing, analysing and manipulating these models can justifiably "affect one's confidence in necessity or impossibility hypotheses" about real-world targets without imposing any similarity, isomorphism, resemblance, etc. requirement between such models and targets (Grüne-Yanoff, 2009, 81). Let us assume, for the sake of argument, that this constitutes an interesting and informative notion of learning.[3] In virtue of *what properties* or *features* can minimal models supposedly prompt such learning?

According to Grüne-Yanoff, "if we are to learn from a model", then this model "must (1) present a relevant possibility that (2) contradicts an impossibility hypothesis that is held with sufficiently high confidence by the potential learners" (2009, 97). In his view, both of these conditions must be satisfied if a model is to prompt learning. More specifically, if condition (2) is not satisfied, then a model "merely shows the possibility of a state that no one believed to be impossible" and therefore "would not affect anyone's confidence level" (ibid., 97). As to condition (1), that the possibility presented by a model is relevant can be established "with reference to natural laws covering this case or to similarity with empirical studies [or to] the credibility of the model" (ibid., 97). Now, neither 'reference to natural laws' nor 'similarity with empirical studies' can be established on the sole basis of information provided by minimal models. For such models are assumed to "lack any similarity, isomorphism or resemblance relation to the world, [be] unconstrained by natural laws or structural identity, and [isolate no] real factors" (Grüne-Yanoff, 2009, 83). Regarding the 'credibility' of models, the following points are worth emphasizing.

Several accounts of models' credibility have been developed in the philosophical literature on scientific modelling (see e.g. de Donato Rodriguez and Zamora Bonilla, 2009, and Mäki, 2009). The proponents of LMM often draw on Sugden's (2000 and 2009) account of credibility on the alleged ground that this account provides "the most relevant arguments supporting the possibility of learning from minimal models"

(see e.g. Grüne-Yanoff, 2009, 81-83) and because adopting the broader characterization would trivialize the issue whether we can learn from minimal models in the sense I explicate in *Section 1*.

[3] I am not concerned here with assessing the merits of this supposition. For present purposes, I note that not all scientific modellers speak of learning in the sense indicated by LMM. In particular, several authors (e.g. Morgan, 1999) regard learning as more akin to acquiring novel information about one's targets than to justifiably changing confidence in specific necessity or impossibility hypotheses about such targets. In what follows, I bracket these definitional concerns and examine what we can learn from minimal models in the sense indicated by LMM. I expand on different ways in which consideration of minimal models can allegedly prompt learning in *Section 3*.

(Grüne-Yanoff, 2009, 89). This does not prevent the proponents of LMM from criticizing specific tenets of Sugden's account. For instance, Grüne-Yanoff doubts Sugden's (2000, 19) claim that credible models can license inductive inferences to general hypotheses about the world by contending that a model's credibility "depends on the intuitions and experiences elicited from particular situations" and that credibility judgements "are highly contingent on the specifics of the situation presented" (2009, 90). Even so, Grüne-Yanoff endorses both Sugden's emphasis on credibility and his characterization of this notion. Let me expand on such characterization.

On Sugden's account, a model is *credible* only if it is logically consistent and the scenarios it envisions 'could be real', in the sense that they "cohere with common intuitions and experience" about the general laws and the causal factors operating in modellers' real-world targets (Sugden, 2000, 26). That a model is credible in this sense indicates that such model is "compatible with what we know, or think we know, about […] the real world" and provides "a description of how the world could be" (Sugden, 2009, 18 and 24). However, it implies neither that the model is true (or approximately true, or probably true, or probably approximately true) nor that it resembles modellers' real world-targets in specific respects. In this regard, the credibility of models closely resembles the credibility of so-called 'realistic' novels. As Sugden puts it, in a realistic novel "the author has to convince us that […] there could be people and places like those in the novel [Nonetheless] the characters and locations are imaginary [and] we do not demand that the events of the novel did happen, or even that they are simplified representations of what really happened" (2000, 25).

Consideration of models that are credible in the sense indicated by Sugden may enable scientific modellers to acquire epistemically informative insights about the *possible worlds* posited by these models. Still, the mere fact that a model is credible in the sense indicated by Sugden falls short of implying that consideration of this model can *per se* prompt learning about *real-world* targets. After all, some credible models point to merely conceivable or logically possible scenarios, mechanisms, etc. that for all modellers know do not characterize the workings or the properties of any real-world target (see e.g. de Donato Rodriguez and Zamora Bonilla, 2009, on economic models assuming perfect information and infinitely divisible quantities, and Weisberg, 2007b, on perpetual motion machines). Moreover, models often make idealizations that interpreted literally cannot be true, and yet are not deemed to lack credibility solely because they make these idealizations (see e.g. Sugden, 2009, on continuity assumptions in economic modelling). These observations, in turn, make it doubtful that consideration of credible models can *per se* prompt learning about real-world targets unless these models provide some information or make some presuppositions concerning such targets.

Faced with these claims, a proponent of LMM might note that credibility judgments "are often elicited solely through consideration of imaginary worlds" (Grüne-Yanoff, 2009, 94). In particular, she might observe that judging a model to be credible does not require one to demonstrate this model's putative similarity, isomorphism, resemblance, etc. to any specific real-world target (ibid., 95). Even so, the point remains that some information or presupposition about real-world targets is needed if one is to *justifiably* take facts about the possible worlds posited by a credible model to stand for putative facts about such targets. In this respect, it is telling that several modellers supplement their credibility judgments with empirical illustrations or 'stories' that purport to explicate how the entities and mechanisms posited by their models may be taken to operate in the real-world situations they target (see e.g. Morgan, 2001). In fact, various authors maintain that grounding informative inferences from the world of a model to the real world requires significant similarities (see e.g. Sugden, 2000) or resemblances (see e.g. Mäki, 2009) between such worlds.

These considerations have far-reaching implications for the merits of LMM. To illustrate this, recall that minimal models are assumed to "lack any similarity, isomorphism or resemblance relation to the world, [be] unconstrained by natural laws or structural identity, and [isolate no] real factors" (Grüne-Yanoff, 2009, 83). As this definition indicates, consideration of a minimal model can *per se* provide neither a priori nor empirical reasons to justifiably infer that what is possible (or necessary) in the worlds envisioned by such model is also possible (or necessary) in the real-world situations investigated by modellers. Therefore, justifiably inferring that the necessity and impossibility results of a minimal model hold also in the real-world situations investigated by modellers requires one to supplement this minimal model with some information or presuppositions concerning those situations. More generally, it appears that consideration of minimal models cannot prompt learning about real-world targets unless one supplements these models with additional information or presuppositions concerning such targets.

I shall expand in the next section on prominent authors' failure to demonstrate that consideration of at least some minimal models does *per se* prompt learning about real-world targets. For the purpose of this section, I consider one illustrative instance of putative learning from minimal models. In recent years, various agent-based models of asset pricing have been proposed to replicate a few so-called 'stylised facts' of finance, i.e. specific statistical properties of financial time series that seemingly persist across a range of time periods and assets (see e.g. LeBaron, 2006, on the fat-tailed distributions of returns and volatility clustering). Some of these models have been claimed to be minimal on the alleged grounds that they do not resemble particular real-world targets, do not aim to reproduce any actual time series in specific time intervals, and so on (see e.g. Casini, 2014, on Arthur et al., 1997, and

Lux and Marchesi, 1999). Those agent-based models of asset pricing purportedly undermine the necessity hypothesis that high price volatility necessarily results from external shocks by demonstrating that such volatility may emerge from the actions of heterogeneous traders even in the absence of external shocks.[4]

Let us suppose, for the sake of argument, that the aforementioned agent-based models of asset pricing may be plausibly regarded as minimal in the sense indicated by LMM. It is hard to see on what grounds these models' possibility results may foster justified changes in confidence in hypotheses about real-world targets unless one provides convincing reasons or evidence (e.g. analogies with biological and physical systems, qualitative match with observed stock series) to think that what is possible (or necessary) in the worlds posited by such models is also possible (or necessary) in the targeted real-world situations. In this respect, one might well rebut that some agent-based models of asset pricing can prompt learning about real-world targets by drawing on *implicit* presuppositions or background information about these targets (see e.g. Alfi et al., 2009, on assumptions concerning the number of agents in the investigated markets). This rebuttal, however, would be of little help to the proponents of LMM. For it relinquishes the claim that these models are minimal in the sense indicated by LMM, and thereby prevents instances of learning from those models from providing support to LMM.

A proponent of LMM might object that it would be exceedingly demanding to require her to specify in virtue of what properties or features minimal models can supposedly prompt learning about real-world targets. After all, the objection would go, a *wide range* of properties and features may prompt learning in different modelling contexts. Furthermore, whether a given property or feature can prompt learning presumably depends on *contextual* factors such as modellers' aims, what information they possess concerning the investigated targets, and so on. These considerations are not without merit, but do not exempt the proponents of LMM from the need to explicate in virtue of what properties or features minimal models can supposedly prompt learning. For the alleged context-dependency of what these properties or features are implies neither that *any* property or feature may enable learning nor that whether specific minimal models possess learning-enabling properties or features is an *arbitrary* matter.

[4] Some authors claim that learning from minimal models may go beyond undermining specific necessity or impossibility hypotheses (see e.g. Casini, 2014, on the alleged provision of mechanistic insights about the examined targets). However, these authors typically adopt less restrictive characterizations of 'minimal model' than the characterization underlying LMM, since they allow that some world-linking relation may hold between what they call 'minimal models' and real-world targets (see e.g. Casini, 2014, 660-662). For present purposes, I focus on the thesis that consideration of models that are minimal in the sense indicated by LMM can justifiably affect confidence in necessity or impossibility hypotheses.

A proponent of LMM might further object that scientific modellers occasionally take their models to prompt learning about real-world targets even in cases where the relation between such models and targets is *uncertain* or *not univocally determined* (see e.g. Grüne-Yanoff, 2009, 83). Still, this objection does not license the claim that consideration of minimal models can *per se* prompt learning about real-world targets. For minimal models are assumed to lack any similarity, isomorphism, resemblance, etc. relation to such targets, and the mere fact that modellers at a given time have not univocally determined what relation holds between a model and their real-world targets falls short of implying that such model is minimal in the sense indicated by LMM (e.g. modellers may subsequently succeed in determining what relation holds between their model and targets).

Similarly, observing that scientific modellers do not generally *specify* what world-linking relation holds between their models and targets (see e.g. Grüne-Yanoff, 2009, 86) would hardly help the proponents of LMM. For this observation by no means implies that models can prompt learning about real-world targets in the absence of world-linking relations. In particular, one can consistently maintain that *different* world-linking relations may enable a model to prompt learning about real-world targets, and yet insist that *some* such relation must hold for the model to prompt learning about those targets (see e.g. footnote no.1; see also Rohwer and Rice, 2013, for a discussion).

At this stage, a proponent of LMM might concede that a literal interpretation of this thesis appears to be implausibly strong.[5] At the same time, she might attempt to identify and substantiate a non-trivial reformulation of LMM. In the points *1.1* and *1.2* below, I consider and rebut two such purported reformulations in turn.

1.1 Qualifying LMM

The first reformulation holds that all LMM was meant to assert is simply that consideration of minimal models, supplemented with *some additional* information or presuppositions concerning real-world

[5] Indeed, one might question whether many scientific models qualify as minimal in the sense indicated by LMM. After all, the thought would be, given a model and a putative target, one can almost invariably find some sort of similarity, isomorphism, resemblance, etc. between such model and target. I am not concerned here with assessing the cogency of this supposition. For the purpose of my critique, it suffices to note that if no model qualified as minimal in the sense indicated by LMM, then LMM would reduce to the vacuous conditional statement that if there were minimal models, then we could learn from such models. In this respect, the challenge for the proponents of LMM is to provide a reformulation of LMM that is neither vacuous in this sense nor trivial in the sense specified in the text (see my remarks on LMM* and LMM** below).

targets, can prompt learning about those targets. Let us call this assertion 'LMM*' so as to distinguish it from the stronger LMM.[6] LMM* seems far less controversial than LMM. Regrettably, LMM* also seems to trivialize the question whether we can learn from minimal models, in the sense that it gives this question an answer that is straightforwardly determinable on the basis of our background knowledge or available evidence. After all, no sensible modellers would deny that consideration of minimal models - supplemented with suitably detailed information and presuppositions about real-world targets - can prompt learning about such targets. Moreover, if LMM* was all the proponents of LMM intended to show, then speaking of 'learning from minimal models' would be quite misleading. For on LMM*, consideration of minimal models is said to prompt learning about real-world targets only in a rather indirect and derivative sense.

Analogous remarks apply to the proposal to regard information from modellers' *context of inquiry* as part of candidate minimal models on the alleged ground that models typically "consist of both a formal structure and an interpretation of this structure" (Grüne-Yanoff, 2009, 84). More specifically, this proposal invites the following two-fold rejoinder. First, if information from modellers' context of inquiry is considered part of a model, then it becomes highly dubious that this model can be plausibly deemed to be minimal in the sense indicated by LMM (e.g. such model may have significant similarity, isomorphism, resemblance, etc. relations to real-world targets). And second, redefining the notion of minimal model so that information from modellers' context of inquiry counts as part of such model makes LMM vulnerable to trivialization in the sense explicated in the previous paragraph.

1.2 Redefining the notion of minimal model

The second reformulation of LMM, which draws on a redefinition of the notion of minimal model embedded in LMM, proceeds as follows. On a literal reading, LMM presupposes a narrow characterization of 'minimal model', according to which only models that *actually lack* any resemblance, similarity, isomorphism, etc. relation to real-world targets count as minimal models. However, one might take the set of minimal models to include both models that actually lack any similarity, isomorphism, resemblance, etc. relation to real-world targets and models for which no relation of similarity, isomorphism, resemblance, etc. to specific real-world targets has been *determined* by modellers (see footnote no.2). Adopting this broader characterization of 'minimal

[6] I am not aware of authors who explicitly relinquish LMM to advocate LMM*. Grüne-Yanoff might be taken to speak in favour of LMM* when he observes (2009, 98) that minimal models do not support hypotheses about particular real-world situations without empirical information about the world (see also Grüne-Yanoff, 2013a).

model', in turn, enables one to defend the claim that we can learn from minimal models (Grüne-Yanoff, personal correspondence).

Does this reasoning point to a plausible and non-trivial reformulation of LMM? It does not seem so. To see this, suppose we take the set of minimal models to include both models that actually lack any similarity, isomorphism, resemblance, etc. relation to real-world targets and models for which no relation of similarity, isomorphism, resemblance, etc. to specific real-world targets has been determined by modellers. Let us call the claim that we can learn from minimal models defined in this broad sense 'LMM**'. At first glance, LMM** might appear to constitute a plausible reformulation of LMM. However, LMM** also seems to trivialize the issue whether we can learn from minimal models, since it merely re-states well-known results in the epistemology of modelling. After all, modellers have reiterated for decades that models can prompt learning in cases where no resemblance, similarity, isomorphism, etc. relation to specific real-world targets has been determined (see e.g. Fisher, 1930, on models of three-sexed sexually reproducing populations, and Weisberg, 2007b, on perpetual motion machines).

Regrettably, several reformulations of LMM would count as trivial in this sense. For instance, consider the hypothetical reformulation of LMM as the thesis that highly idealized models can prompt learning about real-world targets. This reformulation of LMM trivializes the issue whether we can learn from minimal models by redefining the notion of minimal model in such a way that models that have already been shown to prompt learning about real-world targets (see e.g. Mäki, 2005 and 2009, and Sugden, 2000 and 2009) count as 'minimal models'. That is to say, while a literal reading of LMM seemingly fails to withstand scrutiny, the proponents of LMM have hitherto failed to identify and substantiate a non-trivial reformulation of such thesis.

2. Does consideration of *some* minimal models prompt learning?

The critique articulated in *Section 1* challenges the proponents of LMM to specify in virtue of what properties or features minimal models can supposedly prompt learning about real-world targets. In this section, I complement this challenge with two detailed cases studies to argue that consideration of minimal models cannot *per se* prompt learning about such targets. In particular, I aim to demonstrate that LMM falls prey to the following dilemma. On the one hand, truly minimal models - i.e. models that are minimal in the sense indicated by LMM - lack the evidential and epistemic resources to prompt learning about real-world targets. On the other hand, several models that *prima facie* seem minimal in the sense indicated by LMM can prompt learning about real-world targets, but succeed in doing so only if supplemented with additional information or presuppositions about such targets.

In the points *2.1* and *2.2* below, I illustrate this dilemma with regard to two widely cited examples of purportedly minimal models, namely Schelling's (1969, 1971, 1972, and 1978) checkerboard model of residential segregation and Maynard Smith and Parker's (1976) Hawk-Dove model of animal conflict. In each case, I first present the relevant model and explicate in what sense this model is claimed to prompt learning about real-world targets. I then argue that even these widely cited examples of putative learning from minimal models fail to support LMM.[7]

2.1 Schelling's checkerboard model

Schelling's checkerboard model is one of the most cited and discussed works in the philosophical literature on scientific modelling (see e.g. Aydinonat, 2007, Sugden, 2000, and Ylikoski and Aydinonat, 2014). Schelling presents his model as "an abstract exploration of some of the quantitative dynamics of segregating behaviour" which aims to elucidate "what kinds of segregation or integration may result from individual choice" (1971, 147-8). In particular, he purports to demonstrate that observable aggregate phenomena are "compatible with types of 'molecular movement' that do not closely resemble the aggregate outcomes that they determine" (1978, 152). This demonstration, in turn, allegedly vindicates the general principle that individuals' interactions may lead to unintended and unexpected results in a wide range of settings (Schelling, 2006, ch.18).[8]

In Schelling's model, two types of tokens are initially distributed randomly over a checkerboard. These two types of tokens represent two types of individuals and the checkerboard represents a city divided into neighbourhoods. Schelling defines each individual's neighbourhood as the set of grid elements adjacent to the cell occupied by the individual (Moore neighbourhood). Regarding the individuals' types, Schelling contends that his 'ultimate' concern is segregation by 'color' in the United States (1971, 144). At the same time, he makes clear that his analysis "is so abstract that any twofold [exhaustive, and recognizable] distinction could constitute an interpretation - whites and blacks, boys and girls, officers and enlisted men, [etc.]" (ibid., 144).

[7] I focus on these two models because the proponents of LMM have debated over the epistemic import of those models in great detail and have claimed such models to constitute "a good example" of how minimal models can prompt learning about real-world targets (see e.g. Grüne-Yanoff, 2009, 96, on Schelling's model). My considerations hold *mutatis mutandis* for other instances of putative learning from minimal models (see e.g. Batterman and Rice, 2014, on Fisher's model of the 1:1 sex ratio, and Knuuttila, 2009, on Tobin's ultra-Keynesian model).

[8] In his works, Schelling presents several versions of the checkerboard model, which differ in various details (see e.g. Sugden, 2011). I shall expand on these differences whenever they are material to my critique of LMM.

The dynamics of Schelling's model can be explicated as follows. Individuals sequentially choose to either remain in the same place or move to nearby unoccupied cells. Each individual's decision depends on whether her preference to have more than a given proportion of neighbours of the same type as her is satisfied. If an individual's preference about the composition of her neighbourhood is satisfied, then the individual remains where she is. Otherwise, the individual moves to the nearest unoccupied cell where her preference is satisfied. This sequence of decisions continues until all individuals' preferences are satisfied.

In his works, Schelling demonstrates that an *abstract* pattern of segregation - i.e. a pattern of segregation figuring in the possible worlds envisioned by his model - can emerge even from individuals' mild preferences not to be in a minority in the neighbourhood they live in (e.g. Schelling, 1971, assumes that each individual remains in the same place if at least 30% of her neighbours are of the same type as her). After the publication of Schelling's model, several studies (see e.g. Fossett and Dietrich, 2009, Pancs and Vriend, 2007, and Zhang, 2004) have documented the robustness of Schelling's result across a wide range of modifications in his model (e.g. think of changes in neighbourhood sizes, spatial configurations, and individuals' tolerance thresholds for individuals of a different type). In fact, recent studies have shown that segregation may emerge even in cases where individuals strictly prefer to live in diverse communities and be surrounded by agents of a different type (see e.g. Muldoon et al., 2012).

Over the last few years, various proponents of LMM have claimed that Schelling's model provides "a good example" of how minimal models can prompt learning about real-world targets (see e.g. Grüne-Yanoff, 2009, 96). In their view, Schelling's result undermines the belief that segregation is necessarily caused by explicitly racist preferences, thereby leading all those who endorsed such belief to justifiably reduce their confidence in it. This, in turn, constitutes learning in the sense indicated by LMM. Two questions regarding these authors' claims should be carefully distinguished. First, is Schelling's checkerboard model *minimal* in the sense indicated by LMM? And second, does consideration of this model (on the supposition that the model is minimal in such sense) prompt *learning* about real-world targets? Let us address these two questions in turn.

Schelling's checkerboard model does not directly represent an actual segregation process in any real-world city (see e.g. Rohwer and Rice, 2013, 352). In particular, it disregards features that are commonly associated with real-world segregation processes (see e.g. Weisberg, 2012, on interactions across neighbourhoods) and does not include well-known causes of real-world segregation (see e.g. Aydinonat, 2007, on interpersonal welfare differences). Still, it is questionable whether Schelling's model is minimal in the sense indicated by LMM, i.e.

actually "lack[s] any similarity, isomorphism or resemblance relation to the world, [is] unconstrained by natural laws or structural identity, and [does not] isolate any real factors" (Grüne-Yanoff, 2009, 83). Indeed, there are various reasons to doubt that Schelling's model is minimal in this sense. To give one example, the possible worlds envisioned by such model seem to resemble both real-world cities and real-world segregation processes in several respects, ranging from the checkerboard's division into neighbourhoods to the dependency of individuals' actions on whether their preferences are satisfied.

These observations make it dubious that Schelling's model is minimal in the sense indicated by LMM. Even so, a proponent of LMM might insist that *Schelling himself* deemed his model to be minimal in this sense. After all, the thought would be, Schelling "does not justify [his model's assumptions] with reference to the real world at all" (Grüne-Yanoff, 2009, 88). Moreover, "neither similarity, isolation nor conforming to regularity are explicit concerns in Schelling's original paper" (ibid., 88). I am not concerned here with engaging in exegetical debates over Schelling's interpretation of his model. For the purpose of my critique, I note that there are reasons to doubt that Schelling deemed his model to be minimal in the sense indicated by LMM. For instance, Schelling (1971) introduces the checkerboard model after presenting an even more idealized one-dimensional segregation model where coins are arrayed on a line (rather than a grid) and there are no empty spaces, so that when a coin moves, it places itself between two other coins without altering the ordering of any other coin (see e.g. Sugden, 2011). Furthermore, in later works Schelling supplements the checkerboard model with a discussion of real-world segregation processes that hints at putative similarities between his model and those processes (see e.g. Schelling, 1978 and 2006, ch.18).

Let us suppose, for the sake of argument, that Schelling's model is minimal in the sense indicated by LMM. Assume further that Schelling deemed his model to be minimal in this sense. According to some authors, Schelling's demonstration that abstract segregation can emerge even from individuals' mild discriminatory preferences prompts learning about real-world segregation processes in the sense indicated by LMM. Their reasoning goes as follows. When Schelling published his original model, "many people believed that segregation is necessarily [caused by] explicitly racist preferences" (Grüne-Yanoff, 2009, 96). Schelling's demonstration undermines this widely shared belief, thereby leading all those who held such belief to justifiably reduce their confidence in it. This, in turn, constitutes learning in the sense indicated by LMM.

Does this reasoning show that consideration of Schelling's model (on the supposition that this model is minimal) prompts learning about real-world segregation processes? There are at least two reasons to doubt this. First, it is dubious that when Schelling published his original model many people believed that segregation is *necessarily* caused by explicitly racist preferences. And second, Schelling's demonstration that abstract segregation is not necessarily caused by explicitly racist preferences does not *per se* prompt justified changes in confidence in hypotheses about any *real-world* segregation process. Let me expand on these two issues.

When Schelling published his original model, many people presumably believed that residential segregation is *often* - or even *typically* - caused by explicitly racist preferences. This, however, falls short of implying that many (or even some) people held the much stronger belief that segregation is *necessarily* caused by explicitly racist preferences. Moreover, it is dubious that this belief was widely held at the time Schelling published his model. For at that time several factors other than explicitly racist preferences were already well-known possible causes of segregation (see e.g. Schelling, 1969, on interpersonal welfare differences). Indeed, it is questionable whether Schelling's result was altogether surprising for other modellers. For although the degree of segregation that emerges in this model is much higher than individuals' preferences for homogeneity, individuals are still assumed to prefer some degree of homogeneity in their respective neighbourhoods (Muldoon et al., 2012). Taken together, these observations make it doubtful that the necessity hypothesis that is allegedly contradicted by Schelling's model was widely endorsed at the time this model was published.

Let us suppose, for the sake of argument, that when Schelling published his original model, many people believed that segregation is necessarily caused by explicitly racist preferences. Even this would fall short of implying that Schelling's model (on the supposition that this model is minimal) prompts learning in the sense indicated by LMM. To be sure, Schelling's demonstration that individuals' mild discriminatory preferences can foster abstract segregation may prompt a justified change in confidence in hypotheses about the segregation processes figuring in the *possible* worlds envisioned by his model. This demonstration might be epistemically valuable to modellers, but does not imply a justified change in modellers' confidence in hypotheses about any *real-world* segregation process (see *Section 1*). Therefore, such demonstration does not *per se* prompt learning in the sense indicated by LMM.

In fact, it is hard to see on what basis one could justifiably take Schelling's demonstration regarding abstract segregation to prompt learning about real-world segregation processes unless one supplements his model with additional information or presuppositions about such processes (e.g. that individuals live in neighbourhoods, that they are concerned about their neighbourhoods' ethnic composition, etc.). As Sugden puts it, "suppose we read Schelling as claiming that *if* people lived in checkerboard cities, and *if* people came in just two colours, and *if* each person was content provided that at least a third of his neighbours were the same colour as him, and *if*..., and *if*... (going on to list all the properties of the model), *then* cities would be racially segregated. That is not an empirical claim at all: it is a theorem" (2000, 17).

2.2 Maynard Smith and Parker's Hawk-Dove model

As to Maynard Smith and Parker's Hawk-Dove model, the following remarks are in order. Biological modellers often envision highly idealized models to investigate necessity and impossibility hypotheses concerning biological patterns or features observed across heterogeneous real-world populations. One such biological pattern relates to conspecifics' propensity to exercise restraint in combat (rather than fighting to death) when they compete for limited resources such as nesting sites and mating opportunities (see e.g. Maynard Smith and Price, 1973). Since the winners of these conflicts gain resources (e.g. territory and mates) that enable them to transmit genes to future generations at higher frequencies than the losers, adaptationist thinking predicts that individual-level selection would favour fierce physical combat. Hence, the thought would be, group-level selection is required to induce the often-observed pattern of restraint in combat (see e.g. Wynne-Edwards, 1963).

In a pioneering work, Maynard Smith and Price (1973) combine game theoretic methods and computer simulation to investigate the necessity hypothesis that group-level selection is required to induce the often-observed pattern of restraint in combat. Maynard Smith and Parker (1976) build on Maynard Smith and Price's work to demonstrate that *pace* this necessity hypothesis, individual-level selection acting alone can lead to restraint in combat in a wide range of populations. More specifically, Maynard Smith and Parker present a 'Hawk-Dove' game-theoretic model of recurrent, anonymous conflict between pairs of contestants that compete over an indivisible resource and are drawn at random from a large population. The model focuses on situations where the two contestants can rely on some commonly observable asymmetry between their positions to resolve their conflict without having to fight (see e.g. Sugden, 2011).

In the basic Hawk-Dove model, there is a single asymmetry between contestants (e.g. 'discoverer' of the resource and 'latecomer'), which is assumed to correlate neither with their fighting ability nor with the fitness payoffs of winning and losing the contested resource. Moreover, each contestant has just two available strategies, namely escalate until injured or until the opponent retreats (Hawk), and display and then retreat if the opponent escalates (Dove). Three kinds of interactions can occur within the model (Maynard Smith and Parker, 1976). When a Hawk interacts with another Hawk, each contestant has a 50% chance of winning the resource and a 50% chance of being injured. When a Hawk interacts with a Dove, the Hawk wins the resource and the Dove retreats. When a Dove interacts with another Dove, each is equally likely to win the resource without having to fight.[9]

The model has two evolutionarily stable strategies, namely 'escalate if discoverer, retreat if latecomer' and 'retreat if discoverer, escalate if latecomer' (Maynard Smith and Parker, 1976). Each of these evolutionarily stable strategies induces a conventional resolution of the conflict. The idea is that individual-level selection acting alone can lead to restraint in combat in hypothetical situations of animal conflict with common knowledge of an uncorrelated asymmetry and with the Hawk-Dove model's payoff structure (see e.g. Sugden, 2009). This result, in turn, allegedly undermines the necessity hypothesis that group-level selection is required to induce the often-observed real-world pattern of restraint in combat (see e.g. Maynard Smith and Parker, 1976, 171). As with Schelling's checkerboard model, two issues are usefully distinguished. First, is the Hawk-Dove model plausibly regarded as *minimal* in the sense indicated by LMM? And second, does consideration of this model (on the supposition that the model is minimal in such sense) prompt *learning* about real-world targets? Let us consider these two issues in turn.

The Hawk-Dove model makes several idealizations, including infinite population size, asexual reproduction and random pairing of contestants. Moreover, many of the features of the possible worlds posited by such model have no counterparts in any real-world population. For instance, the model implausibly assumes that the payoff structure remains constant across individuals and game iterations (see e.g. Sugden, 2011). Furthermore, the Hawk-Dove model's assumption that the contestants' asymmetry is completely uncorrelated with fighting ability and payoffs is almost invariably violated in real-world populations (see e.g. Maynard Smith and Parker, 1976, 159). Because of these idealizations, the Hawk-Dove model does not accurately represent the selection dynamics of any real-world population (see e.g. Rohwer and Rice, 2013, 341). Yet, in spite of those idealizations, the Hawk-Dove model is not plausibly deemed to be *minimal* in the sense indicated by LMM. For example, the hypothetical populations envisioned by the model resemble real-world populations in their being composed by reproducing organisms whose interactions can critically affect their fitness. Moreover, Maynard Smith and Parker supplement their model with some empirical illustrations that hint at putative similarities between the model's selection dynamics and the selection processes operating in real-world populations.

[9] Maynard Smith and Price (1973) explore more complicated versions of the Hawk-Dove model, where individual contestants can adopt several different strategies and can modify their own behaviour in response to their opponent's behaviour. I focus on the basic Hawk-Dove model because it constitutes a more plausible candidate minimal model than the more complicated versions explored by Maynard Smith and Price (1973).

Let us suppose, for the sake of argument, that the Hawk-Dove model is minimal in the sense indicated by LMM. Does consideration of this model (on the supposition that the model is minimal in such sense) prompt *learning* about real-world targets? It does not seem so. To see this, let us examine the Hawk-Dove model in greater detail. This model demonstrates that individual-level selection acting alone can lead to restraint in combat in hypothetical situations of animal conflict with common knowledge of an uncorrelated asymmetry and with the Hawk-Dove payoff structure (see e.g. Sugden, 2009). This demonstration may prompt a justified change in modellers' confidence in hypotheses about the patterns of restraint in combat figuring in the *possible* worlds posited by Maynard Smith and Parker. However, it does not *per se* constitute learning in the sense indicated by LMM, since it does not imply a justified change in confidence in hypotheses about any *real-world* pattern of restraint in combat. Indeed, it is hard to see how consideration of the Hawk-Dove model (on the supposition that the model is minimal) could prompt learning about real-world populations unless this model is supplemented with additional information or presuppositions about such populations (see point *2.1* above for a similar remark about Schelling's model).

Three parallels between Schelling's checkerboard model and Maynard Smith and Parker's Hawk-Dove model are especially relevant for our appraisal of LMM. First, in both cases a highly idealized model that does not accurately represent any real-world target is employed with the aim to investigate necessity or impossibility hypotheses concerning a range of real-world targets (see e.g. Sugden, 2009). Second, in both cases consideration of the possible worlds envisioned by the examined model enables modellers to demonstrate that a particular process or factor (e.g. mild discriminatory preferences, individual-level selection) can generate the feature or pattern of interest (e.g. segregation processes, restraint in combat) in a range of hypothetical situations (see e.g. Rohwer and Rice, 2013). Finally, and crucially, in both cases calls in support of LMM fall prey to the following dilemma. Either the examined model is minimal in the sense indicated by LMM, in which case it lacks the evidential and epistemic resources to prompt learning about real-world targets. Or the model *prima facie* seems minimal in the sense indicated by LMM, but can prompt learning about real-world targets only if supplemented with additional information or presuppositions about such targets.

3. Minimal models and the epistemology of scientific modelling

Faced with the previous case studies, a proponent of LMM may acknowledge that often-quoted instances of putative learning from minimal models fail to support this thesis. In particular, she might concede that prominent calls in support of LMM are vulnerable to severe objections. Still, she might argue that in spite of these objections, LMM remains a plausible thesis. The idea would be to show that consideration of minimal models can prompt learning about real-world targets even in cases where these models are not supplemented with additional information or presuppositions concerning such targets. In this section, I examine and refute three arguments one may put forward to defend LMM against my critique. I shall consider in turn the argument from how-possibly explanations, the argument from clusters of models, and the argument from heuristic applications. Each of these arguments purports to defend LMM by pointing to a particular way in which consideration of minimal models can allegedly prompt learning about real-world targets. For each argument, I extend the remarks in Fumagalli (2015) and illustrate how my critique of LMM bears on the wider debate on the epistemology of scientific modelling.

3.1 Argument from how-possibly explanations

In the philosophical literature on scientific modelling, a number of authors have contrasted *how-actually* explanations and *how-possibly* explanations (see e.g. Forber, 2010, and Reiner, 1993). This contrast can be explicated as follows. On the one hand, how-actually explanations purport to identify what events or factors in fact cause the occurrence or some specific properties of the investigated phenomena (e.g. what socio-economic factors cause segregation in a particular city, what selection dynamics cause restraint in combat in a specific population). On the other hand, how-possibly explanations merely aim to identify possible causes of those phenomena's occurrence or properties.[10]

The *argument from how-possibly explanations* builds on this distinction between how-actually and how-possibly explanations to put forward the following defence of LMM. Modellers are frequently unable to provide how-actually explanations of the phenomena they investigate, and rely on a menu of how-possibly explanations (see e.g. Ylikoski and Aydinonat, 2014, on economic models; see also Rohwer and Rice, 2013, on biological models). Consideration of minimal models may enable modellers to identify formerly overlooked possible causes of the examined phenomena, thereby extending the set of how-possibly explanations on which they can rely. This contribution, in turn, may lead modellers to justifiably change their confidence in specific

[10] The expression 'how-possibly explanations' was originally used by Dray (1957 and 1968) to indicate explanations that purport to account for how events whose occurrence was previously regarded as impossible could have occurred. Here I employ such expression to designate explanations that aim to identify possible causes of the examined phenomena, irrespective of whether the occurrence of these phenomena was previously regarded as impossible (for another characterization of how-possibly explanations that drops the presumption of impossibility, see e.g. Resnik, 1991; for a recent discussion of how-possibly explanations, see e.g. Bokulich, 2014, Forber, 2012 and Reydon, 2012).

hypotheses about real-world targets. Hence, the argument goes, consideration of minimal models can prompt learning about such targets.

To illustrate how this argument is supposed to substantiate LMM, let us consider again Schelling's checkerboard model of segregation. As we have seen in *Section 2*, this model demonstrates that abstract segregation patterns - i.e. patterns of segregation figuring in the possible worlds envisioned by such model - can emerge even from individuals' mild preferences not to be in a minority in the neighbourhood they live in. Individuals' mild discriminatory preferences are unlikely to be the most prominent cause of real-world segregation processes. For in real-world situations, segregation typically results from other factors such as organized discrimination and economic inequalities between distinct social groups (see e.g. Ylikoski and Aydinonat, 2014). Even so, individuals' mild discriminatory preferences may foster real-world segregation when some other contributing factor (e.g. organized discrimination) is present and might arguably foster segregation even in the absence of other contributing factors. In light of these observations, a proponent of LMM might contend that Schelling's model identifies one previously overlooked possible cause of segregation and thereby extends the set of how-possibly explanations of real-world segregation that are available to modellers (see e.g. Grüne-Yanoff, 2013b). Let us assess the cogency of this contention.[11]

Suppose that Schelling's model extends the set of available how-possibly explanations of segregation by identifying one previously overlooked possible cause of *abstract* segregation. This contribution may be epistemically valuable to modellers (see e.g. Sugden, 2000 and 2009). However, it does not *per se* foster justified changes in confidence in hypotheses concerning *real-world* segregation processes. To be sure, modellers may occasionally be able to demonstrate that the possible cause of abstract segregation identified by Schelling's model can foster segregation also in the real-world situations they investigate (e.g. think of cases where independent studies provide modellers with this information). Still, on the supposition that Schelling's model is minimal, this demonstration would require modellers to supplement such model with information or presuppositions regarding those real-world situations (see *Section 2*). More generally, it remains difficult to see how exactly consideration of a minimal model could enable one to establish that what counts as a possible cause of a phenomenon in the

[11] Rohwer and Rice (2013, 341 and 349) put forward similar remarks concerning Maynard Smith and Parker's (1976) Hawk-Dove model of animal conflict. However, Rohwer and Rice adopt a less restrictive characterization of 'minimal model' than the characterization underlying LMM, since they allow that some world-linking relation may hold between what they call 'minimal models' and real-world targets (see e.g. Rohwer and Rice, 2013, 347). Hence, their calls for the possibility of learning from minimal models do not directly support LMM, and my critique of LMM does not directly apply to such calls (see footnote no.13 for a similar point about Batterman and Rice, 2014).

possible worlds posited by this model is a possible cause of such phenomenon also in real-world situations unless one supplements the examined model with additional information or presuppositions about those situations.

At this stage, a proponent of LMM might conjecture that consideration of minimal models can prompt learning by providing *partial* explanations of the examined real-world phenomena, i.e. by identifying only a subset of the factors that account for the occurrence or specific properties of those phenomena (see e.g. Hempel, 1962). Regrettably, consideration of models that are minimal in the sense indicated by LMM cannot *per se* show that the factors figuring in the associated partial explanations can plausibly account for the occurrence or the properties of the examined real-world phenomena. For these models lack the evidential and epistemic resources to prompt learning about such phenomena. This does not imply that prompting learning about real-world targets requires modellers to accurately represent the behaviour and the properties of these targets. Still, the point remains that *some* world-linking relation must hold between a model and real-world targets if consideration of this model is to prompt learning about such targets. To put it differently, putative partial explanations based solely on minimal models cannot foster justified changes in confidence in hypotheses about real-world targets in the absence of information or presuppositions about such targets.

3.2 Argument from clusters of models

Over the last few decades, most contributions to the philosophical debate on the epistemic import of scientific models have focused on *individual* models and the relationship between individual models and their targets. These contributions aim to explicate in virtue of what properties or features individual models can represent their targets and what conditions these models must satisfy to prompt learning about such targets. In recent years, various authors have criticized this focus on individual models. In particular, some have argued that the epistemic import of scientific models is best understood with reference to *clusters of models* relevant to the modellers' aims (see e.g. Godfrey-Smith, 2006, and Weisberg, 2007a). In their view, scientific modellers frequently rely on clusters of models to learn about their targets, and "one cannot fully appreciate the epistemic import of such models by way of singling out one model from this cluster […] and analyzing it in isolation" (Ylikoski and Aydinonat, 2014, 22).[12]

The *argument from clusters of models* builds on these considerations to

[12] The expression that models 'come into clusters' has been employed in different senses by scientific modellers. Below I use this expression to indicate situations where modellers' predictive and explanatory goals are best achieved by using a combination of structurally dissimilar models.

demonstrate that even if individual minimal models fail to prompt learning about real-world targets, clusters of minimal models can prompt such learning by showing that the implications of specific minimal models can be derived by means of several independent assumptions. To see how consideration of a cluster of minimal models is supposed to prompt learning about real-world targets, let us examine again Schelling's checkerboard model of segregation. As noted in *Section 2*, Schelling's result that an abstract pattern of segregation can emerge from individuals' mild discriminatory preferences holds across a number of modifications in Schelling's model (e.g. changes in neighbourhood sizes, spatial configurations, and individuals' tolerance thresholds for individuals of a different type). This demonstration points to the so-called 'derivational robustness' of models' implications, i.e. the degree to which these implications hold under variations in the assumptions used to derive them (Woodward, 2006). Let me expand on this notion.

Scientific modellers often construct models having different auxiliary assumptions with the aim to determine whether these models' implications hold across dissimilar modelling contexts. This robustness analysis can help modellers to identify which auxiliary assumptions affect their models' implications and determine what difference these assumptions make to those implications. In the philosophical literature on scientific modelling, various authors take the demonstration that a model's implications are derivationally robust to provide some form of epistemic support to such implications (see e.g. Levins, 1966, and Weisberg, 2006). In particular, some note that models' implications frequently rest on unrealistic assumptions and hold that derivational robustness analysis can justifiably increase modellers' confidence in the robust theorems which connect these assumptions to specific modelling results (see e.g. Kuorikoski et al., 2010 and 2012). A proponent of LMM may draw on these contributions to argue that scientific modellers can justifiably increase their confidence in hypotheses about real-world targets by relying on clusters of minimal models.[13]

Let us grant that scientific modellers can justifiably increase their confidence in hypotheses about real-world targets by relying on clusters of models, as opposed to individual models. Even so, appealing to

[13] Batterman and Rice emphasize a similar point when they claim that minimal models can be used to explain why classes of heterogeneous target systems display the same large-scale behavior "by demonstrating that the details that distinguish the model system and [the target] systems are irrelevant" (2014, 350). However, Batterman and Rice adopt a less restrictive characterization of 'minimal model' than the characterization underlying LMM, since they allow that some world-linking relation may hold between what they call 'minimal models' and real-world targets (see e.g. Batterman and Rice, 2014, 355). Hence, their calls for the possibility of learning from minimal models do not directly support LMM, and my critique of LMM does not directly apply to such calls (see footnote no.11 for a similar point about Rohwer and Rice, 2013).

derivational robustness does not *per se* provide a convincing defence of LMM. To illustrate this, imagine facing a situation where modellers rely on a cluster of models that are minimal in the sense indicated by LMM, i.e. a cluster of models each of which "lack[s] any similarity, isomorphism or resemblance relation to the world, [is] unconstrained by natural laws or structural identity, and [does not isolate] any real factors" (Grüne-Yanoff, 2009, 83).

Consideration of a cluster of models that are minimal in this sense may prompt justified changes in modellers' confidence in hypotheses concerning the *possible* worlds posited by these models. Nonetheless, it cannot prompt justified changes in modellers' confidence in hypotheses concerning *real-world* targets unless one supplements at least *some* of these models with additional information or presuppositions about such targets. For consideration of models that are minimal in the sense indicated by LMM cannot *per se* justifiably increase modellers' confidence that the results obtained in the possible worlds posited by such models hold also in the real-world situations they target (see *Section 1*; see also Kuorikoski and Lehtinen, 2009, 127, for a similar remark about results derived from clusters of models that rest on wholly unrealistic assumptions).

3.3 Argument from heuristic applications

The *argument from heuristic applications* purports to defend LMM by contrasting two distinct ways in which consideration of minimal models can putatively prompt learning about real-world targets. The argument can be articulated as follows. Suppose that the proponents of LMM fail to identify plausible cases where consideration of minimal models *directly* prompts justified changes in confidence in hypotheses about real-world targets. This would cast doubt on several calls in support of LMM, but would fall short of undermining LMM. For minimal models have many informative heuristic applications, in the sense that they inspire novel hypotheses about modellers' targets and suggest more precise formulations of former hypotheses about such targets (see e.g. Grüne-Yanoff, 2013b). These heuristic applications, in turn, could *indirectly* foster justified changes in confidence in hypotheses about real-world targets (e.g. consideration of minimal models may lead modellers to form novel beliefs about real-world targets, which in turn challenge specific hypotheses concerning such targets). Hence, the argument goes, consideration of minimal models can prompt learning about real-world targets even in cases where these models are not supplemented with additional information or presuppositions concerning such targets.

There are at least two reasons to doubt that this argument provides a convincing defence of LMM. First, it is often difficult for modellers to demonstrate that a model actually inspires novel hypotheses about their targets and suggests more precise formulations of former hypotheses about such targets (see e.g. Schulz, 2012, on how pointing to trivial reformulations of previous hypotheses does not license the claim that one's model has informative heuristic applications). And second, showing that consideration of minimal models occasionally inspires novel hypotheses about modellers' targets and suggests more precise formulations of former hypotheses about such targets does not *per se* provide a convincing defence of LMM. For neither inspiring novel hypotheses nor suggesting more precise formulations of former hypotheses *per se* constitutes learning in the sense indicated by LMM.[14]

At this stage, a proponent of LMM may well rebut that heuristic applications of minimal models can foster justified changes in confidence in hypotheses concerning the *possible worlds* posited by such models (see e.g. Schlimm, 2009, on how the existence of mathematical models can demonstrate the internal consistency of hypotheses that were formerly deemed to be impossible on a priori grounds). This rebuttal, however, would hardly help the proponents of LMM to substantiate this thesis. For heuristic applications based solely on minimal models cannot prompt justified changes in confidence in hypotheses about *real-world* targets in the absence of information or presuppositions concerning how the possible worlds posited by those models are related to such targets (see Pincock, 2012, 489, for a similar remark regarding mathematical models of biological patterns). To put it differently, heuristic applications based solely on minimal models lack the evidential and epistemic resources to prompt learning about real-world targets unless one supplements the minimal models on which these heuristic applications are based (or those heuristic applications themselves) with additional information or presuppositions concerning such targets.

Conclusion

In the recent literature on the epistemology of scientific modelling, several authors have claimed that highly idealized scientific models can prompt learning about real-world targets. In particular, some have maintained that consideration of minimal models can provide modellers with epistemically informative insights about the possible worlds posited by such models. These two claims have attracted increasing consensus among philosophers of science, but fall short of licensing the stronger 'LMM' thesis that consideration of minimal models can *per se* prompt learning about real-world targets. The proponents of LMM have hitherto failed to provide convincing support to this thesis. Indeed,

[14] As Grüne-Yanoff puts it: "if model use merely contributes to the formulation of [a] hypothesis, then no learning takes place. Learning requires that the model effects justified changes of our confidence in the hypothesis, or if the model itself gave rise to its formulation that it forces us to form a belief about this newly-formulated hypothesis based on consideration of the model itself" (2009, 85).

LMM itself appears to substantially overstate the epistemic import of minimal models even with regard to widely cited examples of putative learning from minimal models. That is to say, *contra* LMM, consideration of minimal models cannot prompt learning about real-world targets unless one supplements these models with additional information or presuppositions concerning such targets.

Acknowledgments

I thank J. McKenzie Alexander, Jesus Zamora Bonilla, Lorenzo Casini, Till Grüne-Yanoff, Olivier Roy, Paul Teller, and Michael Weisberg for their comments on earlier versions of this paper. I also benefited from the observations of audiences at the Munich Center for Mathematical Philosophy, the 6th Models and Simulations Conference at the University of Notre Dame, and the 24th Biennial Meeting of the Philosophy of Science Association in Chicago. An earlier version of part of the paper appeared in the Proceedings of the 24th Biennial Meeting of the Philosophy of Science Association.

REFERENCES

Alfi, V., Cristelli, M., Pietronero, L. and Zaccaria, A. 2009. Minimal agent based model for financial markets I. *Eur. Phys. J*. B, 67, 385–397.

Arthur, W.B., LeBaron, B., Palmer, B., and Taylor, R. 1997. Asset Pricing under Endogenous Expectations in an Artificial Stock Market. In Arthur, W.B. et al. Eds., *Economy as an Evolving Complex System II*, vol.XXVII, 15-44. Reading, MA: Addison-Wesley.

Aydinonat, N.E. 2007. Models, conjectures and exploration: An analysis of Schelling’s checkerboard model of residential segregation. *Journal of Economic Methodology*, 14, 429–454.

Batterman, R. 2002. Asymptotics and the Role of Minimal Models, *British Journal for the Philosophy of Science*, 53, 21-38.

Batterman, R. and Rice, C. 2014. Minimal Model Explanations. *Philosophy of Science*, 81, 349-376.

Bokulich, A. 2014. How the Tiger Bush Got Its Stripes: ‘How Possibly’ vs. ‘How Actually’ Model Explanations. *The Monist*, 97, 321-338.

Cartwright, N. 1999. Capacities. In J. B. Davis, D. W. Hands and U. Mäki (Eds.), *The handbook of economic methodology*, Edward Elgar, 45–48.

Cartwright, N. 2009. If no capacities then no credible worlds. But can models reveal capacities? *Erkenntnis*, *70*, 45–58.

Casini, L. 2014. Not-so-minimal models. Between isolation and imagination. *Philosophy of the Social Sciences*, 44, 646-672.

De Donato Rodriguez, X. and Zamora Bonilla, J. 2009. Credibility, idealisation, and model building: An inferential approach. *Erkenntnis*, 70, 101-118.

Dray, W. 1957. *Laws and explanations in history*. Oxford: Clarendon Press.

Dray, W. 1968. On Explaining How-Possibly. *The Monist*, 52, 390-407.

Fisher, R.A. 1930. *The Genetical Theory of Natural Selection*. Oxford: Clarendon Press.

Forber, P. 2010. Confirmation and explaining how possible. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 41, 32–40.

Forber, P. 2012. Conjecture and explanation: a reply to Reydon. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, 298–301.

Fossett, M. and Dietrich, D. 2009. Effects of city size, shape, and form, and neighborhood size and shape in agent-based models of residential segregation: are Schelling-style preference effects robust? *Environment and Planning B: Planning and Design*, 36(1), 149-169.

French, S. 2003. A model-theoretic account of representation (or, I don't know much about art…but I know it involves isomorphism). *Philosophy of Science*, 70, 1472–1483.

Fumagalli, R. 2015. No Learning from Minimal Models. *Philosophy of Science*. Proceedings of the 24th Biennial Meeting of the Philosophy of Science Association.

Giere, R.N. 2004. How models are used to represent reality. *Philosophy of Science*, 71, 742–752.

Godfrey-Smith, P. 2006. The strategy of model-based science. *Biology and Philosophy*, 21, 725-740.

Goldenfeld, N. 1992. *Lectures on phase transitions and the renormalization group*. Frontiers in physics, Vol.85. Reading (MA): Addison-Wesley.

Grüne-Yanoff, T. 2009. Learning from Minimal Economic Models, *Erkenntnis*, 70, 81-99.

Grüne-Yanoff, T. 2013a. Appraising Models Nonrepresentationally. *Philosophy of Science*, 80, 850-861.

Grüne-Yanoff, T. 2013b. Genuineness resolved: a reply to Reiss' purported paradox. *Journal of Economic Methodology*, 20, 255-261.

Hausman, D.M. 1992. *The inexact and separate science of economics*. Cambridge University Press.

Hempel, C.G. 1962. Explanation in Science and History. In *Frontiers of Science and Philosophy.* Colodny, R. Ed. 9-19. University of Pittsburgh Press.

Knuuttila, T. 2009. Isolating representations vs. credible constructions? Economic modelling in theory and practice. *Erkenntnis*, 70, 59-80.

Kuorikoski, J. and Lehtinen, A. 2009. Incredible Worlds, Credible Results. *Erkenntnis*, 70, 119-131.

Kuorikoski, J., Lehtinen, A. and Marchionni, C. 2010. Economic modelling as robustness analysis. *British Journal for the Philosophy of Science*, 61, 541-567.

Kuorikoski, J., Lehtinen, A. and Marchionni, C. 2012. Robustness analysis disclaimer: please read the manual before use! *Biology and Philosophy*, 27, 891-902.

LeBaron, B. 2006. Agent-based Computational Finance. In Tesfatsion, L. and Judd K.L. Ed., *Handbook of Computational Economics. Agent-based Computational Economics*, 2, 1187-1233. Elsevier.

Levins, R. 1966. The strategy of model building in population biology. *American Scientist*, 54, 421-431.

Lux, T. and Marchesi, M. 1999. Scaling and criticality in a stochastic multi-agent model of a financial market**.** *Nature*, 397, 498-500.

Mäki, U. 1992. On the method of isolation in economics. *Poznan´ Studies in the Philosophy of Science and the Humanities*, 26, 316–351.

Mäki, U. 2005. Models are experiments, experiments are models. *Journal of Economic Methodology*, 12, 303–315.

Mäki, U. 2009. MISSing the world. Models as isolations and credible surrogate systems. *Erkenntnis*, 70, 29-43.

Maynard Smith, J. and Parker, G.A. 1976. The Logic of Asymmetric Contests. *Animal Behavior*, 24, 159-175.

Maynard Smith, J. and Price, G.R. 1973. The Logic of Animal Conflict. *Nature*, 246, 15-18.

Morgan, M.S. 1999. Learning from Models. In Morgan and Morrison (Eds.), *Models as mediators: Perspectives on natural and social science*. Cambridge University Press, 347–388.

Morgan, M.S. 2001. Models, stories and the economic world. *Journal of Economic Methodology*, 8, 361–397.

Muldoon, R., Smith, T. and Weisberg, M. 2012. Segregation That No One Seeks. *Philosophy of Science*, 79, 38–62.

Pancs, R. and Vriend, N.J. 2007. Schelling's spatial proximity model of segregation revisited. *Journal of Public Economics*, 91, 1-24.

Pincock, C. 2012. Mathematical Models of Biological Patterns: Lessons from Hamilton's Selfish Herd. *Biology and Philosophy*, 27, 481-496.

Reiner, R. 1993. Necessary Conditions and Explaining How-Possibly. *Philosophical Quarterly*, 44, 58-69.

Resnik, D.B. 1991. How-Possibly Explanations in Biology. *Acta Biotheoretica*, 39, 141-149.

Reydon, T. 2012. How-possibly explanations as genuine explanations and helpful heuristics: a comment on Forber. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, 302-310.

Rohwer, Y. and Rice, C. 2013. Hypothetical Pattern Idealization and Explanatory Models. *Philosophy of Science*, 80, 334-355.

Schelling, T.C. 1969. Models of segregation. *American Economic Review*, 59, 488–493.

Schelling, T.C. 1971. Dynamic models of segregation. *Journal of Mathematical Sociology*, 1, 143–186.

Schelling, T.C. 1972. The process of residential segregation: Neighbourhood tipping. In Pascal, A. (Ed.), *Racial discrimination in economic life*, 157–185. Lexington Books.

Schelling, T.C. 1978. *Micromotives and Macrobehavior*. Norton, New York.

Schelling, T.C. 2006. *Strategies of commitment and other essays*. Cambridge: Harvard University Press.

Schlimm, D. 2009. Learning from the existence of models: on psychic machines, tortoises, and computer simulations. *Synthese*, 169, 521–538.

Schulz, A.W. 2012. Beyond the Hype: The Value of Evolutionary Theorizing in Economics. *Philosophy of the Social Sciences*, 20, 1–27.

Suarez, M. 2004. An inferential conception of scientific representation. *Philosophy of Science*, 71, S767–779.

Sugden, R. 2000. Credible worlds: The status of theoretical models in economics. *Journal of Economic Methodology*, 7, 1-31.

Sugden, R. 2009. Credible Worlds, Capacities and Mechanisms, *Erkenntnis*, 70, 3–27.

Sugden, R. 2011. Explanations in search of observations, *Biology and Philosophy*, 26, 717-736.

Sugden, R. 2013. How fictional accounts can explain. *Journal of Economic Methodology*, 20, 237-243.

Swoyer, C. 1991. Structural Representation and Surrogative Reasoning. *Synthese*, 87, 449–508.

Teller, P. 2001. Twilight of the Perfect Model Model. *Erkenntnis*, 55, 393-415.

Van Fraassen, B.C. 1980. *The scientific image*. Oxford: Clarendon Press.

Weisberg, M. 2006. Robustness analysis. *Philosophy of Science*, 73, 730–742.

Weisberg, M. 2007a. Three Kinds of Idealization. *The Journal of Philosophy*, 104, 639-659.

Weisberg, M. 2007b. Who is a modeler? *British Journal for the Philosophy of Science*, 58, 207–233.

Weisberg, M. 2012. Getting Serious about Similarity. *Philosophy of Science*, 79, 785–794.

Woodward, J. 2006. Some varieties of robustness. *Journal of Economic Methodology*, 13, 219–240.

Wynne-Edwards, V.C. 1963. Intergroup selection in the evolution of social systems. *Nature*, 200, 623-626.

Ylikoski, P. and Aydinonat, N.E. 2014. Understanding with Theoretical Models. *Journal of Economic Methodology*, 21, 19-36.

Zhang, J. 2004. Residential segregation in an all-integrationist world. *Journal of Economic Behavior & Organization*, 54, 533-550.